NUMERICAL ANALYSIS PROJECT
MANUSCRIPT NA-89-07

JUNE 1989

Iterative Methods for Cyclically Reduced Non-Self-Adjoint Linear Systems II

by

Howard C. Elman
and
Gene Golub

NUMERICAL ANALYSIS PROJECT
COMPUTER SCIENCE DEPARTMENT
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

UMIACS-TR-89-45 June1989
CS-TR-2238

Iterative Methods for Cyclically Reduced Non-Self-Adjoint Linear Systems II

Howard C. Elman[1]
Department of Computer Science
and Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742
Gene H. Golub[2]
Department of Computer Science
Stanford University
Stanford, CA 94305

Abstract

We perform an analytic and experimental study of line iterative methods for solving linear systems arising from finite difference discretizations of non-self-adjoint elliptic partial differential equations on two-dimensional domains. The methods consist of performing one step of cyclic reduction, followed by solution of the resulting reduced system by line relaxation. We augment previous analyses of one-line methods, and we derive a new convergence analysis for two-line methods, showing that both classes of methods are highly effective for solving the convection-diffusion equation. In addition, we compare the experimental performance of several variants of these methods, and we show that the methods can be implemented efficiently on parallel architectures.

Abbreviated Title. Iterative Methods for Reduced Systems.

Key words: Linear systems, reduced system, iterative methods, convection-diffusion, non-self-adjoint.

AMS(MOS) subject classification. Primary: 65F10, 65N20. Secondary: 15A06.

[1]The work of this author was supported by the National Science Foundation under grants DMS-8607478, CCR-8818340, and ASC-8958544, and by the U. S. Army R-h Office under grant DAAL-0389-K-0016.

[2]The work of this author was supported by the National Science Foundation under grant DCR-8412314, the Simon Guggenheim Memorial Foundation, and the University of Maryland's Institute for Advanced Computer Studies, whose support is gratefully acknowledged.

1. Introduction.

We consider iterative methods for solving linear systems of the type that arise from two-cyclic discretizations of t-dimensional elliptic partial differential equations. Such systems can be ordered using a *red-black ordering* so that they have the form

$$\begin{pmatrix} D & C \\ E & F \end{pmatrix} \begin{pmatrix} u^{(r)} \\ u^{(b)} \end{pmatrix} = \begin{pmatrix} v^{(r)} \\ v^{(b)} \end{pmatrix} \tag{1.1}$$

where D and F are diagonal matrices. If block elimination is used to decouple the "red" points $u^{(r)}$ from the "black" points $u^{(b)}$, the result *is* a *reduced system*

$$[F - ED^{-1}C]u^{(b)} = v^{(b)} - ED^{-1}v^{(r)}. \tag{1.2}$$

Let

$$S = F - ED^{-1}C, \qquad s = v^{(b)} - ED^{-1}v^{(r)}. \tag{1.3}$$

In [4], we showed that the coefficient matrix S is also sparse, and we analyzed a class of iterative methods for solving (1.2) when (1.1) comes from a finite-difference discretization of the constant coefficient convection-diffusion equation

$$\mathcal{A}u = -\Delta u + \sigma u_x + \tau u_y = f \tag{1.4}$$

with Dirichlet boundary conditions. In particular, we showed that although S is typically nonsymmetric, it can be symmetrized in a wide variety of circumstances. The symmetrized form was used to analyze the convergence properties of a splitting operator based on a block Jacobi splitting of S, using a one-line *ordering* of the underlying grid.

In this paper, we refine and augment the analysis of [4]. We show that if (1.1) is derived from the convection-diffusion equation (1.4), then the reduced system is itself a discretization of the differential equation. We consider a variety of orderings of the rows and columns of S and examine their effects on the convergence of iterative methods for solving (1.2), and on implementation. In particular, we present several variants of the one-line ordering of [4] based on red-black and *toroidal* groupings of unknowns. In addition, we present an analysis of *two-line* ordering strategies for solving (1.2); such orderings have been studied for self-adjoint problems in [6], [12]. In all of these cases, the reduced matrices have block Property A so that Young's analysis of iterative methods [18] is applicable. We use this analysis to determine the convergence properties of block Jacobi, Gauss-Seidel and successive overrelaxation (SOR) methods for solving the discrete convection-diffusion equation, in terms of discrete cell Reynolds numbers $ah/2$ and $\tau h/2$. In addition, we present the results of numerical experiments showing some effects of ordering strategies not revealed by the analysis. Together, the analytic and numerical results show that the two types of orderings are very effective for solving (1.4), with the two-line orderings somewhat more effective than the one-line orderings. The variants of the methods based on red-black orderings of the reduced system are typically slightly slower (in terms of iteration counts), but they can be implemented more efficiently on parallel architectures.

An outline of the paper is as follows. In §2, we describe two discretization schemes for (1.4), and we present an analysis of the truncation error associated with taking the reduced system as an approximation of (1.4). In §3, we present several variants of the one-line ordering for the unknowns of (1.2), and we show how the results of [4] are used to derive a convergence analysis of all the associated one-line iterative methods when the linear system comes from (1.4). In §4, we present the two-line orderings and the convergence analysis of the corresponding two-line iterative methods applied to (1.2). In §5, we outline an analysis due to Parter [12] and Parter and Steuerwalt [14] that complements our results in the limiting case $h \to 0$. In §6, we describe numerical experiments that confirm and supplement the convergence analysis, including tests in which the block iterative methods, with various orderings, are used to solve a set of nonsymmetric problems derived from (1.4). Finally, in §7 we draw some conclusions.

2. The convection-diffusion equation and the reduced system.

Consider the two-dimensional convection-diffusion equation (1.4), posed on the unit square $\Omega \in (0,1) \times (0,1)$ with Dirichlet boundary conditions u = g on $\partial\Omega$. Discretization by a five-point finite difference operator leads to a linear system

$$A\,u - v$$

where u now denotes a vector in a finite dimensional space. We discretize on a uniform n x n grid using standard second order differences for the Laplacian [17], [18], and either centered or upwind differences for the first derivatives. With u ordered lexicographically in the natural ordering as $(u_{1,1}, u_{2,1}, \ldots, u_{n,n})^T$, the coefficient matrix has the form

$$A = tri\,[\,A_{j,j-1},\, A_{jj},\, A_{j,j+1}\,]. \tag{2.1}$$

Here, $tri\,[\,X_{j,j-1}, X_{jj}, X_{j,j+1}\,]$ is the (block) tridiagonal matrix whose j'th row contains $X_{j,j-1}$ X_{jj} and $X_{j,j+1}$ on its subdiagonal, diagonal and superdiagonal, respectively. The subdiagonal of the first row and the superdiagonal of the last row are not defined. The *subscripts will be omitted when there is no ambiguity. The entries of (2.1) are

$$A_{j,j-1} = bI, \qquad A_{jj} = tri\,[c, a, d], \qquad A_{j,j+1} = eI,$$

where I is the identity matrix, a, b, c, d and e depend on the discretization, and all blocks are of order n. Let $h = 1/(n+1)$. After scaling by h^2, the matrix entries are given by

$$a = 4, \quad b = -(1+\delta), \quad c = -(1+\gamma),$$
$$d = -(1-\gamma), \quad e = -(1-\delta),$$

for the centered difference scheme, where $\gamma = \sigma h/2$ and $\delta = \tau h/2$; and

$$a = 4 + 2(\gamma + \delta), \quad b = -(1+2\delta), \quad c = -(1+2\gamma),$$
$$d = -1, \quad e = -1,$$

for the upwind scheme. At the (i, j) grid point, the right hand side satisfies $v_{ij} = h^2 f_{ij}$ where $f_{ij} \equiv f(ih, jh)$.

In [4], we showed that the reduced matrix S is a *skewed* nine-point operator. At all grid points except those bordering $\partial\Omega$, the computational molecule has the form (after scaling by a) given in Fig. 2.1. For grid points next to $\partial\Omega$, the diagonal entries of S (center point of the computational molecule) are different. These values are

(2.2)

$a^2 - 2be - cd$	for points with one horizontal and two vertical neighbors in the original grid
$a^2 - be - 2cd$	for points with one vertical and two horizontal neighbors
$a^2 - be - cd$	for points with just two neighbors.

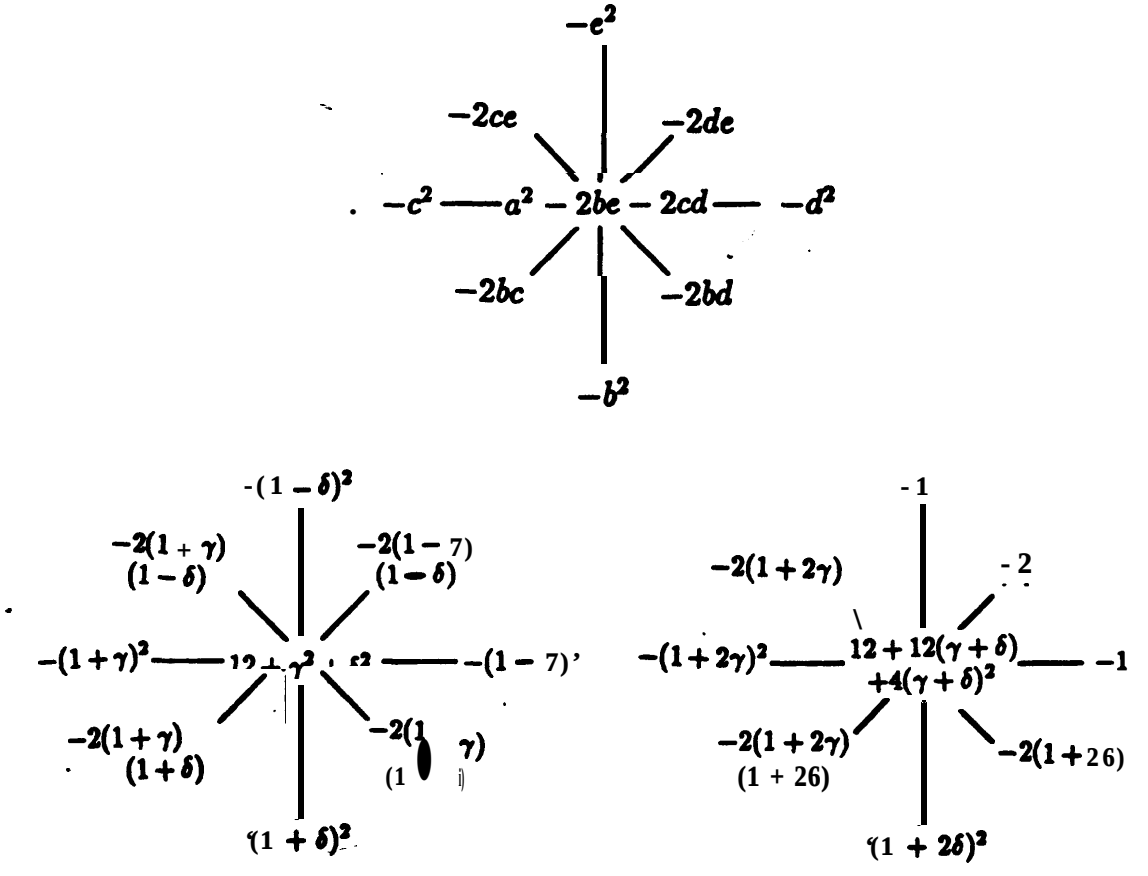

Fig. 2.1: Computational molecules for the reduced system. Top: general case. Bottom left: centered differences. Bottom right: upwind differences.

Suppose centered differences are used to discretize the first derivative terms. At the (i, j) grid point, the discrete-operator, satisfies $\frac{1}{h^2}[Au]_{ij} = [\mathcal{A}u]_{ij} + O(h^2)$, *i.e.* the truncation error of the discretization is of order h^2. The following result shows that the

reduced system (1.2) can also be viewed as a discretization of (1.4) with truncation error of order h^2. When (1.2) arises from the centered difference discretization of (1.4), let $\tilde{S}$ and $\tilde{s}$ denote the reduced matrix and right hand side resulting from multiplying the reduced system by a (= 4).

THEOREM 1. *For $2 \le i, j \le n-1$, the reduced operator $\tilde{S}$ satisfies satisfies*

$$\frac{1}{8h^2}[\tilde{S}u]_{ij} = -\left[(1+\frac{\sigma h^2}{8})u_{xx}+(1+\frac{\tau h^2}{8})u_{yy}\right] + \sigma u_x + \tau u_y + O(h^2),$$

and the reduced right hand side $\tilde{s}$ satisfies

$$\frac{1}{8h^2}\tilde{s}_{ij} = f_{ij} + O(h^2).$$

Proof. The proof follows directly by taking the first five terms of the Taylor series for each of the quantities $u_{i\pm2,j}$, $u_{i,j\pm2}$ and $u_{i\pm1,j\pm1}$, expanded about u_{ij}. Multiplying each entry of $\tilde{S}$ (e.g. from Fig. 2.1) by the appropriate expanded value of u and summing the coefficients for each partial derivative gives

$$\begin{aligned}[\tilde{S}u]_{ij} = 8h^2 \Big[&\sigma u_x + \tau u_y - (1+\tfrac{\sigma h^2}{8})u_{xx} - (1+\tfrac{\tau h^2}{8})u_{yy} - \tfrac{\sigma\tau}{4}h^2 u_{xy} \\ &+ \tfrac{5}{12}\sigma h^2 u_{xxx} + \tfrac{1}{4}\tau h^2 u_{xxy} + \tfrac{1}{4}\sigma h^2 u_{xyy} + \tfrac{5}{12}\tau h^2 u_{yyy} \\ &- \tfrac{5}{24}h^2 u_{xxxx} - \tfrac{1}{4}h^2 u_{xxyy} - \tfrac{5}{24}h^2 u_{yyyy} + O(h^4)\Big].\end{aligned}$$

The reduced right hand side is given by

$$\tilde{s}_{ij} = 4v_{ij} + (1+\delta)v_{i,j-1} + (1+\gamma)v_{i-1,j} + (1-\gamma)v_{i+1,j} + (1-\delta)v_{i,j+1}.$$

Using the fact that $v_{ij} = h^2 f_{ij}$ for all (i, j) and expanding $f_{i,j\pm1}$ and $f_{i\pm1,j}$ in Taylor series about f_{ij} gives

$$\tilde{s}_{ij} = 8h^2 f_{ij} - h^4(-\Delta f + \sigma f_x + \tau f_y) + O(h^5). \qquad \square$$

The expression for $[\tilde{S}u]_{ij}$ in this proof was computed by hand and checked using MACSYMA [9]. The perturbation of the Laplacian (which is also of order h^2) can be thought of as an addition of artificial viscosity, see [16]. A similar analysis shows that the reduced system for the upwind scheme approximates (1.4) with truncation error $O(h)$.

In the following, we use the symbols S and s to represent the reduced matrix and right hand side, respectively, *after* scaling by the diagonal entry a. Our analysis of iterative methods for solving the reduced system (1.2) is based on the fact that in some circumstances, S can be symmetrized by a real diagonal similarity transformation.

THEOREM 2. *There exists a real diagonal matrix Q such that $Q^{-1}SQ$ is symmetric if the product bcde is positive.*

See [4] for a proof. For the centered difference scheme, $be = 1-\delta^2$ and $cd = 1-\gamma^2$, so that S is symmetrizable if both $|\gamma| < 1$ and $|\delta| < 1$ or if both $|\gamma| > 1$ and $|\delta| > 1$. For the

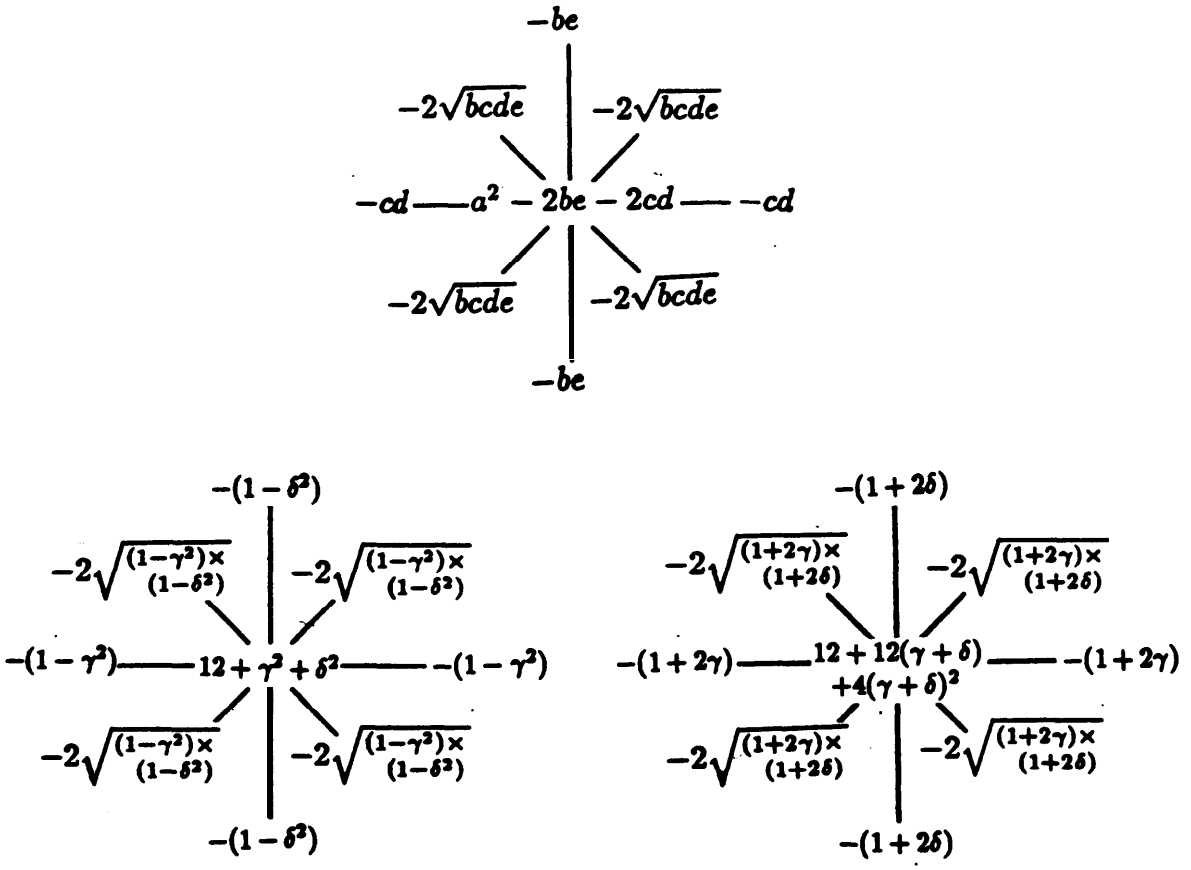

Fig. 2.2: Computational molecules for the symmetrized reduced system. Top: general case. Bottom left: centered differences. Bottom right: upwind differences.

upwind scheme, S is symmetrizable for all nonnegative γ and 6. Computational molecules for the symmetrized matrix are shown in Fig. 2.2.

3. One-he orderings.

The performance of iterative methods for solving (1.2) depends on the ordering of the underlying grid. In this section, we describe and analyze several *one-line* orderings, in which grid points are grouped by diagonal lines oriented at a 45° angle with the horizontal and vertical axes. For the purpose of discussion, we fix the orientation to be along the NW-SE direction. We consider four orderings.

In the *natural one-line ordering,* the $n-1$ diagonal lines are numbered starting from one comer (e.g. the SW) from 1 to $n-1$, and individual points are numbered from bottom to top along the lines. An example for $n=7$ is shown in the left side of Fig. 3.1 where the line indices are shown outside $\partial\Omega$. The corresponding matrix S is block tridiagonal. In the *red-black one-line* ordering, the lines with odd indices from the natural ordering are ordered first, followed by those with even indices. The individual grid points are renumbered to

be consistent with this reordering. An example for $n = 7$ is shown in the right side of Fig. 3.1. Here, the reduced matrix has the form of the coefficient matrix of (1.1), where the block diagonal matrices D and F consist of uncoupled tridiagonal blocks.'

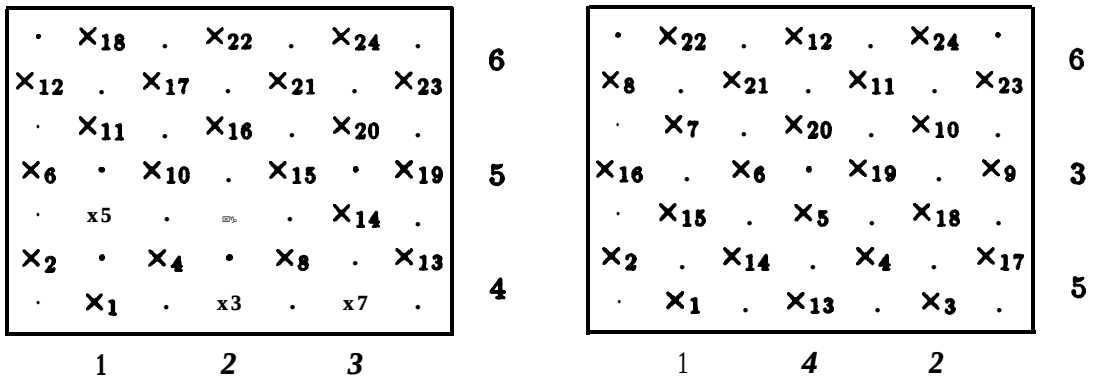

Fig. 3.1: The reduced grid derived from a 7 × 7 grid, with natural one-line (left) and red-black one-line (right) orderings.

Note that the individual lines in the reduced grid, and therefore the associated tridiagonal matrices, vary in size. For the natural ordering, the lines have sizes

$$\begin{cases} 2,\ 4,\ \ldots,\ n-1,\ n-1,\ \ldots,\ 4,\ 2 & \text{for odd } n \\ 2,\ 4,\ \ldots,\ n-2,\ n,\ n-2,\ \ldots,\ 4,\ 2 & \text{for even } n. \end{cases}$$

The other two orderings are defined so that lines of less than maximal size are paired up to form sets of fixed size. This will be of use on parallel architectures (see §6). In the *torus one-line* ordering, each line of less than maximal size from one corner of the grid is followed by the line from the opposite comer that would be obtained by continuing the grid periodically; these pairs of lines then are organized as in the natural ordering. For example, for odd n, the first four lines are the one in the SW comer containing 2 mesh points, followed by the line closest to the NE corner containing $n - 3$ points, the line of size 4 in the SW comer and then the line of size $n - 5$ closest to the NE comer. The ordering for $n = 7$ is shown in the left side of Fig. 3.2. Thus, the reduced grid can be grouped together into $\lceil n/2 \rceil$ sets consisting of either one or two lines, each containing a total of $n - 1$ mesh points. For even n, the analogue produces $n/2$ sets of points, each of size n.

We define the fourth ordering in terms of these fixed sized sets. Suppose first that they are listed consecutively according to their appearance in the torus ordering. For example, for the grid on the left side of Fig. 3.2, the listing is

$$\{1,2\},\ \{3,4\},\ \{5\},\ \{6\},$$

where these integers are those outside the domain in the left side of Fig. 3.2. Now let this listing be permuted in alternating fashion,

$$\{1,2\},\ \{5\},\{3,4\},\ \{6\},$$

and assign indices of increasing value to these sets. As above, let the grid point indices be assigned so that they are consistent with this ordering of lines. We refer to the result as the *alternating torus one-line* ordering. An example is shown in the right side of Fig. 3.2. This ordering is well-defined for all n, but we will show below that it is most useful when $\lceil n/2 \rceil$ is even, where it corresponds to a red-black ordering.

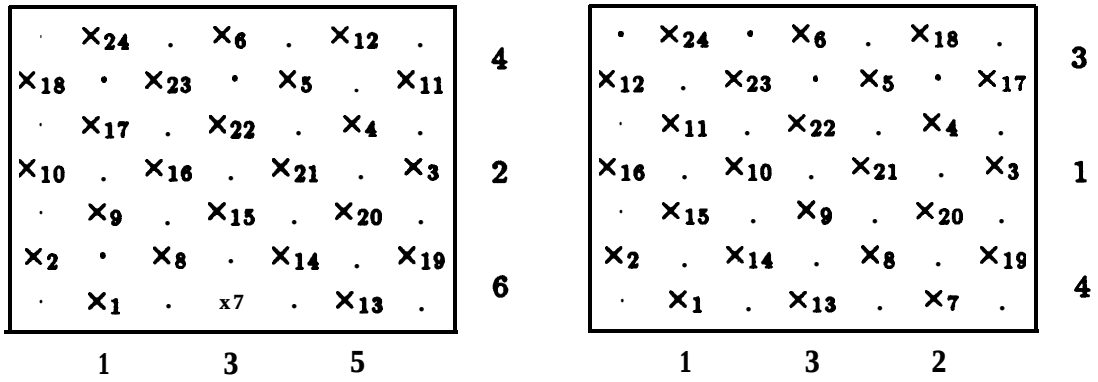

Fig. 3.2: The reduced grid derived from a 7 x 7 grid, with torus one-line (left) and alternating torus one-line (right) orderings.

For all four orderings, the reduced matrix has the form

$$S = D - C, \tag{3.1}$$

where D is a block diagonal matrix whose individual blocks are tridiagonal matrices. Consider the block Jacobi iterative method

$$u_{k+1}^{(b)} = Bu_k^{(b)} + D^{-1}s,$$

where $B = D^{-1}C$ is the block Jacobi iteration matrix. The standard measure of the effectiveness of this method is the spectral radius $\rho(B)$; the iteration is *convergent* provided $\rho(B) < 1$, and convergence is more rapid if $\rho(B)$ is closer to 0 [17] (cf. §6). We have left unspecified the particular ordering determining (3.1); the iteration matrices for the various ordering strategies are all similar to one another, so that $\rho(B)$ is independent of ordering. In [4], we derived bounds on $\rho(B)$ for the version of B arising from the natural ordering. The results are summarized as follows; see [4] for proofs.

THEOREM 3. *For the one-line orderings, if* $be > 0$ *and* $cd > 0$, *then*

$$\rho(B) \le \frac{2(\sqrt{be} + \sqrt{cd})^2}{a^2 - 2(\sqrt{be} + \sqrt{cd})^2 + 4\sqrt{bcde}\,(1 - \cos(\pi h))}.$$

If $be < 0$, $cd < 0$, *then*

$$\rho(B) \le \frac{\max(4\sqrt{bcde}, 2\sqrt{bcde} + |be|, 2\sqrt{bcde} + |cd|, |be| + |cd|) + 2(|be| + |cd|)}{a^2 + 2(\sqrt{|be|} - \sqrt{|cd|})^2 + 4\sqrt{bcde}\,(1 - \cos(\pi h))}.$$

whenever $a^2/2 + (\sqrt{-cd} - \sqrt{-be})^2 - 2\sqrt{bcde} \geq 0$.

COROLLARY 1. For *the centered difference scheme, if* $|\gamma| \leq 1$ *and* $|\delta| < 1$, then *the one-line Jacobi iteration matrices satisfy*

$$\rho(B) \leq \frac{(\sqrt{1-\gamma^2} + \sqrt{1-\delta^2})^2}{8 - (\sqrt{1-\gamma^2} + \sqrt{1-\delta^2})^2 + 2\sqrt{(1-\gamma^2)(1-\delta^2)}(1-\cos(\pi h))}.$$

If $|\gamma| > 1$, $|\delta| > 1$ *and* $\sqrt{(\gamma^2-1)(\delta^2-1)} \leq 4$, *then*

$$\rho(B) \leq \frac{\frac{1}{2}\mu(\gamma,\delta) + \gamma^2 - 1 + \delta^2 - 1}{8 + (\sqrt{\gamma^2-1} - \sqrt{\delta^2-1})^2 + 2\sqrt{(\gamma^2-1)(\delta^2-1)}(1-\cos(\pi h))},$$

where

$$\mu(\gamma,\delta) = \max\,(4\sqrt{(\gamma^2-1)(\delta^2-1)},\ 2\sqrt{(\gamma^2-1)(\delta^2-1)} + \gamma^2 - 1,\ 2\sqrt{(\gamma^2-1)(\delta^2-1)} + \delta^2 - 1,\ \gamma^2 - 1 + \delta^2 - 1).$$

For the upwind difference scheme,

$$\rho(B) \leq \frac{(\sqrt{1+2\gamma} + \sqrt{1+2\delta})^2}{2(2+\gamma+\delta)^2 - (\sqrt{1+2\gamma} + \sqrt{1+2\delta})^2 + 2\sqrt{(1+2\gamma)(1+2\delta)}(1-\cos(\pi h))}.$$

We now show that Young's analysis of relaxation methods also applies to these splittings. Let $C = L + U$, where L and U are strictly lower triangular and upper triangular, and let $\mathcal{L}_\omega = (D-\omega L)^{-1}[(1-\omega)D+\omega U]$ denote the block SOR iteration matrix. Recall the definition of block-consistent orderings from [18]: 8 block matrix $M = M_{ij}, \{1 \leq i, j \leq m\}$ is block-consistently ordered if the integers $1,\ldots,m$ can be partitioned into disjoint sets $\{S_k\}_{k=1}^{t}$ such that if $M_{ij} \neq 0$, then $i \in S_k$ implies $j \in S_{k-1}$ for $j < i$, and $j \in S_{k+1}$ for $j > i$. We have the following result:

LEMMA 1. *For the natural, red-black and torus orderings, the reduced matrix is block-consistently ordered for all* $n > 0$. *For the alternating torus ordering, the reduced matrix is block-consistently ordered if and only if* $\lceil n/2 \rceil$ *is even.*

Proof. The coefficient matrix for the natural ordering is block tridiagonal; the analysis for this ordering and its red-black analogue is classical, see [18]. In discussing the torus and alternating torus orderings, it will be convenient to refer to the line indices of the natural ordering (i.e. from the left side of Fig. 3.1). Let ψ be a mapping of these indices to those of the torus ordering? Then $S_k = \{\psi(k)\}$ determines a consistent ordering. For the alternating torus ordering, note that its block structure is different than for the other orderings, because pairs of lines are grouped together: using the indices of the natural ordering, lines j and $\lceil n/2 \rceil + j$ *are coalesced* into one set. If $\lceil n/2 \rceil$ is *even,* then j and $\lceil n/2 \rceil$ + j have the same parity, and the ordinary red-black coloring of lines determines a

[1] For exampk, for Figs. 3.1-3.2, $\psi(1)=1, \psi(2)=3, \psi(3)=5$, etc. It is possible to derive a precise expression for ψ, but we do not believe it adds insight.

red-black coloring of the alternating torus ordering. A consistent ordering is determined by the partitioning

$$\mathcal{S}_1 = \{\{1, \lceil n/2 \rceil + 1\}, \{3, \lceil n/2 \rceil + 3\}, \ldots\}, \quad \mathcal{S}_2 = \{\{2, \lceil n/2 \rceil + 2\}, \{4, \lceil n/2 \rceil + 4\}, \ldots\}.$$

If $\lceil n/2 \rceil$ is odd, then lines 1 and $\lceil n/2 \rceil + 1$ have the same color, since they comprise one set, but line $\lceil n/2 \rceil$ must share this same color, since (proceeding form the SW comer) alternating lines are assigned opposite colors (see Fig. 3.3). As a result, the alternating torus ordering does not have block Property A, and it therefore cannot be consistently ordered (see [18], §5.4). □

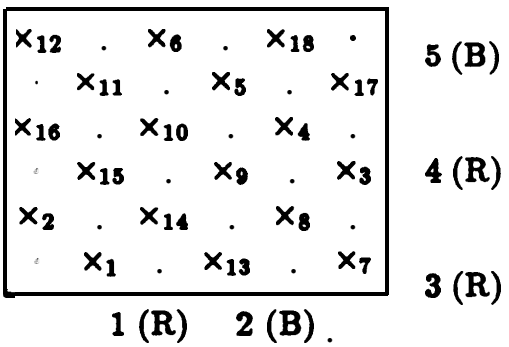

Fig. 3.3: The alternating torus ordering for odd $\lceil n/2 \rceil$. Line indices correspond to the natural one-line ordering. Terms in parentheses indicate that the associated matrix does not have block Property A.

THEOREM 4. *The eigenvalues $\{\mu\}$ of B and $\{\lambda\}$ of $\mathcal{L}_\omega$ are related by*

$$(\lambda + \omega - 1)^2 = \omega^2 \mu^2 \lambda. \tag{3.2}$$

Moreover, if $p(B) < 1$ and either $be > 0$ and $cd > 0$ holds or $be < 0$ and $cd < 0$ holds, then the choice

$$\omega^* = \frac{2}{1 + \sqrt{1 - \rho(B)^2}} \tag{3.3}$$

minimizes $\rho(\mathcal{L}_\omega)$ with respect to ω, and $\rho(\mathcal{L}_{\omega^*}) = \omega^* - 1$.

Proof. The first assertion follows directly from [18], Chapter 14, Theorem 3.4. For the second assertion, it was shown in [4] that if either condition on *be* and cd holds, then $\hat{D}$ is a symmetric positive definite M-matrix. Consequently, all eigenvalues of $\hat{B}$, whence those of B, are real. Therefore, the choice of optimal SOR parameter follows from [18], *$5.2* and §14.3. □

Remarks. When $be > 0$ and $cd > 0$, a sufficient condition to ensure that $\rho(B) < 1$ is

$$a^2 \leq 4(\sqrt{be} + \sqrt{cd})^2, \tag{3.4}$$

which holds for the two difference schemes. In this case, experiments described in [4] indicate that the bounds of Corollary 1 of Theorem 3 for $|\gamma|, |\delta| < 1$ are good indicators of spectral radii. The bound for $|\gamma|, |\delta| > 1$ of Corollary 1 does not always guarantee that $\rho(B) < 1$. However, experimental evidence and Fourier analysis [4] suggest that the smaller bound

$$\rho(B) \leq \frac{(\sqrt{\gamma^2 - 1} + \sqrt{\delta^2 - 1})^2}{8 + (\sqrt{\gamma^2 - 1} + \sqrt{\delta^2 - 1})^2}$$

applies in this case, and this bound is always less than one. Finally, the results of [5], [10] imply that the Chebyshev semi-iterative method applied to the reduced system, with preconditioning by the block diagonal D, has the same asymptotic convergence behavior as the block SOR method with $w = \omega^*$.

4. Two-line orderings.

An alternative to the ordering strategies of the previous section is to group the points of the reduced grid by *pairs* of horizontal or vertical lines. Such *two-line* orderings also result in matrices that have block Property A. Examples with horizontal lines, for $n = 6$, *axe* shown in Fig. 4.1. The left side of the figure shows *a natural two-line ordering,* and the right side shows *a red-black two-line* ordering. In the following, we perform an analysis of two-line orderings for the case of horizontal lines. We use the natural ordering to motivate the analysis; as above, the results also apply to the red-black ordering.

The reduced matrix S for the natural two-line ordering has block tridiagonal form

$$S = tri\,[\, S_{j,j-1}\; S_{jj},\; S_{j,j+1}\,].$$

Within the line pairs, points are ordered from left to right (as in Fig. 4.1), so that the submatrices on the block diagonal are banded. For even n, the block diagonal consists of $n/2$ uncoupled pentadiagonal matrices of order n of the form

$$S_{jj} = \begin{pmatrix} * & -2bd & -d^2 & & \\ -2ce & * & -2de & -d^2 & \\ -c^2 & -2bc & * & -2bd & -d^2 \\ & c^2 & -2ce & * & -2de \\ & & & & \ddots \end{pmatrix},$$

$1 \leq j \leq n/2$. Here, "*" is defined as in the center point of Fig. 2.1, or by (2.2) for points next to $\partial\Omega$. The off-diagonal blocks have the irregular tridiagonal form

$$S_{j,j-1} = -\begin{pmatrix} b^2 & & & & & \\ 2bc & b^2 & 2bd & & & \\ & & b^2 & & & \\ & & 2bc & b^2 & 2bd & \\ & & & & b^2 & \\ & & & & & \ddots \end{pmatrix}, \quad S_{j,j+1} = -\begin{pmatrix} e^2 & 2de & & & & \\ & e^2 & & & & \\ & 2ce & e^2 & 2de & & \\ & & & e^2 & & \\ & & & 2ce & e^2 & 2de \\ & & & & & \ddots \end{pmatrix}.$$

For odd n, the last ($\lceil n/2 \rceil$'th) row and column have slightly different form, in which the last diagonal block is the tridiagonal matrix of order $\lfloor n/2 \rfloor$,

$$tri\,[\,-c^2\,*,\,-d^2\,], \tag{4.1}$$

and the neighboring off-diagonal blocks are adjusted in an analogous manner.

Let D now denote the block diagonal matrix defined by $D_j = S_{jj}$, and let S = D – C denote the two-line Jacobi splitting. Consider the two-line Jacobi iteration

$$u_{k+1}^{(b)} = Bu_k^{(b)} + D^{-1}s,$$

for solving (1.2), where $B = D^{-1}C$. Convergence again depends on $\rho(B)$. Let $\hat{S} = Q^{-1}SQ$ represent the symmetrized reduced matrix when it exists, and let $\hat{D} = Q^{-1}DQ$ and $\hat{C} = Q^{-1}CQ$. We first bound $\rho(B)$ in the case where $be > 0$ and $cd > 0$, i.e. for the centered difference scheme when $|\gamma| < 1$ and $|\delta| < 1$, and for the upwind scheme. The analysis essentially consists of the following two results, which bound the minimum eigenvalue of $\hat{D}$ and maximum eigenvalue of $\hat{C}$. These will then be combined to bound $p(B)$.

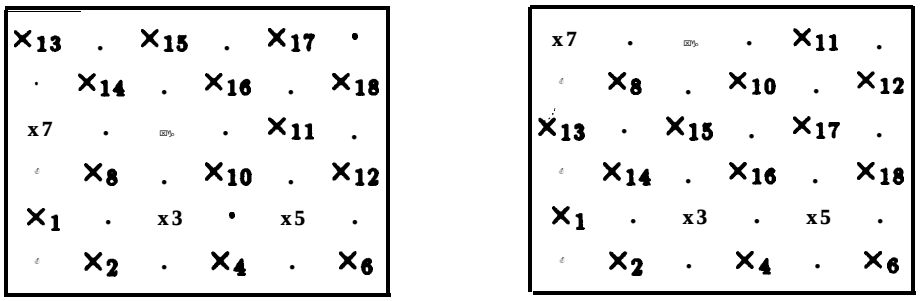

Fig. 4.1: The reduced grid derived from a 6 x 6 grid, with natural two-line (left) and red-black two-line (right) orderings.

LEMMA *2. When be > 0 and cd > 0, the minimum eigenvalue of the symmetrized two-line block diagonal matrix $\hat{D}$ is bounded below by*

$$a^2 - 2(\sqrt{cd} + \sqrt{be})^2 - 2cd + 4\sqrt{bcde}\,(1 - \cos\pi h) + 4cd\,(1 - \cos^2\pi h).$$

Proof. Examination of Figs. 2.2 and 4.1 reveals that all of the matrices on the block diagonal of $\hat{D}$, except the first and last, are identical pentadiagonal matrices of order n. They have the form

$$P \equiv \begin{pmatrix} * & -2\sqrt{bcde} & -cd & & \\ -2\sqrt{bcde} & * & -2\sqrt{bcde} & -cd & \\ -cd & -2\sqrt{bcde} & * & -2\sqrt{bcde} & -cd \\ & -cd & -2\sqrt{bcde} & * & -2\sqrt{bcde} \\ & & -cd & -2\sqrt{bcde} & * \\ & & & \ddots & \ddots \end{pmatrix}, \tag{4.2}$$

where "*" equals $a^2 - 2\&e - 2cd$ except in the first and last entry, where it is $a^2 - 2be - cd$. If P' denotes either the first block, or for even n, the last block of $\hat{D}$, then by (2.2), we have $P' \geq P$ with inequality only on the diagonal. Hence, $\lambda_{min}(P') \geq \lambda_{min}(P)$. A straightforward argument also shows that for all small h, the minimum eigenvalue of $\hat{D}$ does not correspond to an eigenvalue of the tridiagonal matrix (4.1). Hence, it suffices to bound $\lambda_{min}(P)$ below.

For this, let T_n denote the tridiagonal matrix $tri\,[1, 0, 1]$ of order n. Then T_n^2 is a pentadiagonal matrix with O's on the first subdiagonal and superdiagonal, 1's on the second subdiagonal and superdiagonal, and 2's in all diagonal entries except the first and last, where the values are 1. Then we have

$$P = (a^2 - 2be)\, I_n - 2\sqrt{bcde}\, T_n - cd\, T_n^2, \tag{4.3}$$

where I_n is the identity matrix of order n. But the eigenvalues of T_n are $(2\cos(j\pi h)\}_{j=1}^n$, so that those of P are $\{a^2 - 2be - 4\sqrt{bcde}\cos(j\pi h) - 4cd\cos^2(j\pi h)\}_{j=1}^n$. The minimum corresponds to the choice j = 1. □

LEMMA 3. *The maximum eigenvalue of the symmetrized two-line block off-diagonal matrix $\hat{C}$ is bounded by*

$$2\,|be|\cos 2\pi h + 4\sqrt{bcde}\cos \pi h + o(h^2).$$

Proof. Assume n is even; modifications to the argument for odd n are straightforward. Let $\mathcal{R}$ denote the block tridiagonal matrix *tri* [R, 0, R], with $m = n/2$ block rows, where $R = be\, I_n$. Let $\mathcal{V}$ denote the block tridiagonal matrix $tri\,[V^T, 0, V]$, of the same order, where

$$V = \begin{pmatrix} 0 & v & & & \\ & 0 & & & \\ v & 0 & v & & \\ & & 0 & & \\ & & v & 0 & v \\ & & & & \ddots \end{pmatrix}$$

and $v = 2\sqrt{bcde}$. Then $\hat{C} = \mathcal{R} + Y$. Since $\hat{C}$ is symmetric, we have

$$\rho(\hat{C}) = \|\hat{C}\|_2 \leq \|\mathcal{R}\|_2 + \|\mathcal{V}\|_2.$$

To bound $\|\mathcal{R}\|_2$, note that $\mathcal{R}$ is similar to the block diagonal matrix

$$be\ diag\{T_m, \ldots, T_m\}$$

with n block rows, so that its eigenvalues are $\left\{2be\cos\frac{j\pi}{m+1}\right\}_{j=1}^m$. Hence

$$\|\mathcal{R}\|_2 = \rho(\mathcal{R}) = 2|be|\cos\left(\frac{2\pi h}{1+h}\right). \tag{4.4}$$

For $\|\mathcal{V}\|_2$, we have

$$\|\mathcal{V}\|_2 = \|\mathcal{V}^T\mathcal{V}\|_2^{1/2} = \|\mathcal{V}^2\|_2^{1/2}.$$

$\mathcal{V}^2$ is the block pentadiagonal matrix

$$\begin{pmatrix} VV^T & 0 & V^2 & & & \\ 0 & V^TV+VV^T & 0 & V^2 & & \\ (V^T)^2 & 0 & V^TV+VV^T & 0 & V^2 & \\ & & & \ddots & & \\ & & (V^T)^2 & 0 & V^TV+VV^T & 0 \\ & & & (V^T)^2 & 0 & V^TV \end{pmatrix}$$

But $V^2 = 0$, so that in fact $\mathcal{V}^2$ is a block diagonal matrix, and we need only bound the spectral radii of $VV^T, V^TV + VV^T$ and V^TV. We have

$$VV^T = \begin{pmatrix} v^2 & 0 & v^2 & & & \\ 0 & 0 & 0 & 0 & & \\ v^2 & 0 & 2v^2 & 0 & v^2 & \\ & 0 & 0 & 0 & 0 & 0 \\ & & & & \ddots & \\ & & & v^2 & 0 & 2v^2 & 0 \\ & & & & 0 & 0 & 0 \end{pmatrix}, \quad V^TV = \begin{pmatrix} 0 & 0 & 0 & & & \\ 0 & 2v^2 & 0 & v^2 & & \\ 0 & 0 & 0 & 0 & 0 & \\ & & & \ddots & & \\ & & v^2 & 0 & 2v^2 & 0 & v^2 \\ & & & 0 & 0 & 0 & 0 \\ & & & & v^2 & 0 & v^2 \end{pmatrix}$$

Consequently,

$$V^TV + VV^T = \begin{pmatrix} v^2 & 0 & v^2 & & & \\ 0 & 2v^2 & 0 & v^2 & & \\ v^2 & 0 & 2v^2 & 0 & v^2 & \\ & & & \ddots & & \\ & & v^2 & 0 & 2v^2 & 0 \\ & & & v^2 & 0 & v^2 \end{pmatrix} = v^2T_n^2.$$

Thus,

$$\rho(V^TV + VV^T) = v^2[\rho(T_n)]^2 = 16\,bcde\cos^2\pi h. \tag{4.5}$$

Moreover, by permuting VV^T and V^TV so that nonzero entries lie on a tridiagonal band in the upper left comer, we find that

$$\rho(V^TV) = \rho(VV^T) < v^2\rho(2I_m + T_m) = 16\,bcde\cos^2\left(\frac{\pi h}{1+h}\right). \tag{4.6}$$

Bounds *(4.5)* and *(4.6)* are essentially the same as $h \to 0$, which gives the asymptotic result

$$\|\mathcal{V}\|_2 \le 4\sqrt{bcde}\cos\pi h + o(h^2). \tag{4.7}$$

The conclusion then follows from (4.4) and (4.7). □

The idea used in both these proofs of squaring the tridiagonal matrix T_n to generate a pentadiagonal matrix appears in [13], for analyzing line iterative methods applied to discrete biharmonic problems. A simpler argument than the proof of Lemma 3, based on Gerschgorin's theorem, gives the weaker bound $\rho(\hat{C}) \leq 2|be| + 4\sqrt{bcde}$. This bound is close to the result of Lemma 3, but it is less useful for asymptotic analysis as $h \to 0$.

THEOREM 5. *When $be > 0$ and $cd > 0$, the spectral radius of the two-line Jacobi iteration matrix is bounded by*

$$\frac{2\,be\cos 2\pi h + 4\sqrt{bcde}\cos \pi h}{a^2 - 2(\sqrt{cd} + \sqrt{be})^2 - 2cd + 4\sqrt{bcde}\,(1 - \cos \pi h) + 4cd\,(1 - \cos^2 \pi h)} + o(h^2).$$

Proof. Using the similarity transformation $D^{-1}C = Q\hat{D}^{-1}\hat{C}Q^{-1}$, we have

$$\rho(D^{-1}C) = \rho(\hat{D}^{-1}\hat{C}) \leq \|\hat{D}^{-1}\|_2 \|\hat{C}\|_2 = \frac{\rho(\hat{C})}{\lambda_{min}(\hat{D})},$$

where the last equality follows from the symmetry of $\hat{D}$ and $\hat{C}$. The result then follows immediately from Lemmas 2 and 3. □

Substitution of particular values ofa-e gives the following bounds for the two difference schemes under consideration.

COROLLARY 2. *For the centered difference scheme, if $|\gamma| < 1$ and $|\delta| < 1$, then the spectral radius of the two-line block Jacobi iteration matrix for the reduced system is bounded by*

$$\frac{(1 - \delta^2)\cos 2\pi h + 2\sqrt{(1 - \gamma^2)(1 - \delta^2)}\cos \pi h}{8 - (\sqrt{1 - \gamma^2} + \sqrt{1 - \delta^2})^2 - (1 - \gamma^2) + 2\sqrt{(1 - \gamma^2)(1 - \delta^2)}\,(1 - \cos \pi h) + 2(1 - \gamma^2)\,(1 - \cos^2 \pi h)} + o(h^2).$$

For the upwind difference scheme, the spectral radius is bounded by

$$\frac{(1 + 2\delta)\cos 2\pi h + 2\sqrt{(1 + 2\gamma)(1 + 2\delta)}\cos \pi h}{2(2 + \gamma + \delta)^2 - (\sqrt{1 + 2\gamma} + \sqrt{1 + 2\delta})^2 - (1 + 2\gamma) + 2\sqrt{(1 + 2\gamma)(1 + 2\delta)}\,(1 - \cos \pi h) + 2(1 + 2\gamma)(1 - \cos^2 \pi h)} + o(h^2).$$

If (3.4) holds, then the bounds of Theorem 5 are smaller than those of Theorem 3 for the one-line orderings. Consequently, the two-line bounds of Corollary 2 are smaller than the one-line bound8 of Corollary 1.

Now consider the case $be < 0$ and $cd < 0$, which corresponds to the centered difference scheme when $|\gamma| > 1$ and $|\delta| > 1$. To bound p(B), we require an alternative to Lemma 2. Consider the case of odd n.[2] *Let* P be as *in* (4.2), where "$*$" now represents

$$\min(a^2 - be - 2cd, a^2 - 2be - cd) \quad (= \min(13 + 2\gamma^2 + \delta^2, 13 + \gamma^2 + 2\delta^2)). \tag{4.8}$$

[2] In this case, only the first two terms of (2.2) occur. For even n, a somewhat weaker bound can be derived by replacing "$*$" with $a^2 - be - cd$.

For any pentadiagonal matrix P' on the block diagonal of $\hat{D}$, the diagonal entries of P' are greater than those of P, so that $\lambda_{min}(P') \geq \lambda_{min}(P)$. If (4.8) is minimized by $a^2 - be - 2cd$ (i.e. $\gamma^2 \leq \delta^2$), then P satisfies

$$P = (a^2 - be)\, I_n - 2\sqrt{bcde}\, T_n - cd\, T_n^2,$$

(This differs from (4.3) in the coefficient of I_n.) Consequently, all eigenvalues of P have the form

$$a^2 - be - 4\sqrt{bcde}\cos\theta - 4cd\cos^2\theta,$$

for $\theta \in (0, \pi)$. By elementary calculus, we find that this expression is minimized at $\theta = \arccos\left(\frac{1}{2}\sqrt{\left|\frac{be}{cd}\right|}\right)$. The minimum value, a^2, is a lower bound for $\lambda_{min}(\hat{D})$. If (4.8) is minimized by $a^2 - 2be - cd$ $(\gamma^2 \geq \delta^2)$, then

$$P = (a^2 - 2be + cd)\, I_n - 2\sqrt{bcde}\, T_n - cd\, T_n^2.$$

The same argument shows that its minimum eigenvalue is bounded below by $a^2 - be + cd$. As above, these bounds for $\lambda_{min}(P)$ are *smaller* than the minimum eigenvalue derived from (4.1). Combining these observations with Lemma 3, we have the following result.

THEOREM 6. *For $be < 0$ and $cd < 0$ and even n, the spectral radius of the two-line Jacobi iteration matrix is bounded by*

$$\begin{cases} \dfrac{2|be|\cos 2\pi h + 4\sqrt{bcde}\cos\pi h}{a^2} & \text{when (4.8) is minimized by } a^2 - be - 2cd, \\ \dfrac{2|be|\cos 2\pi h + 4\sqrt{bcde}\cos\pi h}{a^2 - be + cd} & \text{when (4.8) is minimized by } a^2 - 2be - cd. \end{cases}$$

For the centered difference discretization when $|\gamma| > 1$ and $|\delta| > 1$, the bounds are

$$\begin{cases} \dfrac{(\delta^2 - 1)\cos 2\pi h + 2\sqrt{(\gamma^2 - 1)(\delta^2 - 1)}\cos\pi h}{8} & \text{for } \gamma^2 < \delta^2, \\ \dfrac{(\delta^2 - 1)\cos 2\pi h + 2\sqrt{(\gamma^2 - 1)(\delta^2 - 1)}\cos\pi h}{8 + \frac{1}{2}(\delta^2 - \gamma^2)} & \text{for } \gamma^2 \geq \delta^2. \end{cases}$$

As we show in §6, the bounds from Theorem 5 and Corollary 2 appear to be tight, whereas the results of Theorem 6 are pessimistic.

Finally, the analysis of [4] implies that for both difference schemes, when $be > 0$ and $cd > 0$, the block pentadiagonal matrix $\hat{D}$ is a symmetric positive definite M-matrix. Hence, we have the following result for the two-line SOR iteration matrix $\mathcal{L}_\omega$.

COROLLARY 3. *For the two-line orderings, the eigenvalues $\{\mu\}$ of B and $\{\lambda\}$ of $\mathcal{L}_\omega$ are related by (3.2). For the two difference schemes under consideration, if $be > 0$ and $cd > 0$ holds, then (3.3) minimizes $\rho(\mathcal{L}_\omega)$ with respect to ω, and $\rho(\mathcal{L}_{\omega^*}) = \omega^* - 1$.*

5. Asymptotic analysis.

In this section, we outline the results of Parter [12] and Parter and Steuerwalt [14] that reveal asymptotic convergence rates as $h \to 0$ for fixed σ and τ in (1.4). (See also [11].) We emphasize that we are only filling in some minor details; all the analysis is contained in [12], [14]. Assume that S is a matrix such that S/h^2 is a discrete approximation to d with truncation error $o(1)$ at all mesh points of Ω not next to the boundary, and O(1) at points next to $\partial\Omega$. Let $S = D - C$ be a splitting. The following result is proved in [14]:

THEOREM 7. *Suppose the following conditions hold for all small h:*

(PS1) $\rho(D^{-1}C) < 1$.

(PS2) $\rho(D^{-1}C)$ *is an eigenvalue of* $D^{-1}C$.

(PS3) $\|C\|_2$ *is bounded independent of h.*

(PS4) *There is a smooth function q satisfying* $q(x, y) \geq q_0 > 0$ *on* $\bar{\Omega}$, *such that*

$$(Cu, v) = (qu, v) + E \tag{5.1}$$

where in (5.1), *q refers to the vector of mesh values, und* $E = he_1(u, v) + h^2 e_2(u, v)$ *depends on* σ *und* τ.

Then as $h \to 0$, $\rho(D^{-1}C) = 1 - \Lambda_0 h^2 + o(h^2)$, *whete* Λ_0 *is the smallest eigenvalue of the problem*

$$\mathrm{Au} = Aqu \quad in\ \Omega, \qquad u = 0\ on\ \partial\Omega. \tag{5.2}$$

In assumption (PS4), e_1 is a function of first order differences in u and u and e_2 is a function of second order differences; see [14] for a more precise statement.

By Theorem 1, the reduced matrix is an appropriate approximation to A. Condition (PSI) has been established in §3 and §4. For both the one-line and two-line Jacobi splittings, condition (PS2) follows from the Perron-Frobenius theory, using the fact that D is an M-matrix for all small enough h [4]. Condition (PS3) follows from Lemma 3. Thus, it remains to determine q for condition (PS4). Much of [12] and [14] is concerned with how to do this. In particular, §7 of [12] and §9 of [14] imply that $q = 1$ for the one-line Jacobi splitting for the reduced system and $q = 3/4$ for the two-line Jacobi splitting. It is straightforward to verify that the eigenfunctions and eigenvalues of (5.2) for $q = 1$ are

$$u^{(j,k)} = e^{\sigma x/2}\sin(j\pi x)\, e^{\tau y/2}\sin(k\pi y), \qquad \Lambda_{jk} = \frac{\sigma^2}{4} + \frac{\tau^2}{4} + (j^2 + k^2)\pi^2,$$

for integers j, $k \geq 1$. The minimum eigenvalue is $\Lambda_0 = \frac{\sigma^2}{4} + \frac{\tau^2}{4} + 2\pi^2$. Hence, we have the following asymptotic result (which applies for both difference schema):

COROLLARY 4. *The spectral radii of the block Jacobi iteration matrices for the one-line orderings of the reduced system are bounded by*

$$1 - \left(\frac{\sigma^2}{4} + \frac{\tau^2}{4} + 2\pi^2\right) h^2 + o(h^2),$$

and the spectral radii of the block Jacobi iteration matrices for the two-line orderings are bounded by

$$1 - \left(\frac{\sigma^2}{3} + \frac{\tau^2}{3} + \frac{8}{3}\pi^2\right)h^2 + o(h^2).$$

For large σ and τ (and small enough h), these bounds are essentially of the form $1 - c(\gamma^2 + \delta^2)$.

The analyses of §3 and §4 give asymptotic bounds of

$$1 - \left(\frac{\sigma^2}{4} + \frac{\tau^2}{4} + \frac{\pi^2}{4}\right)h^2, \quad 1 - \left(\frac{\sigma^2}{3} + \frac{\tau^2}{3} + 2\pi^2\right)h^2,$$

for the one-line and two-line block Jacobi iteration matrices, respectively. These results agree with those of Corollary 4 except in the coefficient of π^2. They are pessimistic because the numerators and denominators come from separate bounds, and (for the one-line case) because Gerschgorin's theorem is used for the numerator. However, it may be more important to know the spectral radius in the nonasymptotic regime, i.e. for particular values of γ and δ not close to zero. The numerical experiments of §6 below indicate that the bounds of §3 and §4 are good indicators of spectral radii in such cases.

Note that smaller values of q *in* Theorem 7 produce smaller spectral radii. The analysis of [14] shows that for the l-line Jacobi splitting of the unreduced system (which gives rise to methods comparable in cost to both methods considered here for the reduced system), $q = 2$. Thus, the asymptotic value of the spectral radius is smaller for the reduced system. An alternative proof of this fact, derived from regular splitting arguments (which are less dependent on asymptotics) is given in [6]. This observation is in agreement with results on spectral radii in [4]. Thus, asymptotic convergence behavior will be worse for the full system.

6. Numerical experiments and implementation.

In this section, we present the results of numerical experiments that confirm and supplement the analysis of §§3 – 4. For the two-line ordering, we compare the bounds on spectral radii of iteration matrices with computed spectral radii, and for all the orderings considered, we examine the performance of the Gauss-Seidel and SOR methods for solving the reduced system arising from the centered difference discretization of the convection-diffusion equation. Except where indicated, all computations were performed on a VAX-8600 in double precision Fortran. The reduced matrices were computed using PCGPAK [15]. All spectral radii were computed using the QZ algorithm in EISPACK [7], [8].

6.1. Spectral radii for the two-line methods.

Tables 6.1 – 6.3 show the computed values of the spectral radii of the Gauss-Seidel iteration matrices for the two-line orderings, for three values of h and different choices of the parameters γ and δ. In addition, the last column of each table shows the asymptotic limits (as $h \to 0$) of the bounds on these spectral radii, when such a bound exists. For $|\gamma|$,

γ	$h = 1/8$	$h = 1/16$	$h = 1/32$	Asymptotic Bound
.2	.42	.74	.86	.90
.4	.33	.55	.63	.66
.6	.22	.34	.38	.40
.8	.11	.16	.18	.19
1.0	.01	.02	.06[3]	.02
1.2	.03	.04	.04	
1.4	.05	.06	.06	
1.6	.06	.06	.07	
1.8	.07	.07	.07	
2.0	.07	.07	.07	
3.0	.07	.07	.07	

Table 6.1: Spectral radii and bounds for the two-line Gauss-Seidel iteration matrices, centered differences, $\delta = 0$.

$|\delta| \leq 1$, these quantities are the squares of the limiting values from Corollary 2, where the values for γ or $\delta = 1$ are the limits as $\gamma, \delta \to 1$. For Table 6.3 when $|\gamma| > 1$, we use Theorem 7. As in [4], the experimental results show that the bounds are good approximations to the limits as $h \to 0$ when $|\gamma| \leq 1$ and $|\delta| \leq 1$, and the bounds for $|\gamma|$, $|\delta| > 1$ are pessimistic. For values of γ and δ where the analysis does not apply, the computed spectral radii are very close to zero. Note that the asymptotic results are expressed in a nonstandard way. In contrast to the analysis of §5, γ and δ are *fixed* here as $h \to 0$, so the continuous problem (1.4) is varying.

6.2. Performance of the block iterative methods.

Figs. 6.1 – 6.3 summarize the performance of the block iterative methods for solving various examples of the discrete convection-diffusion equation (1.4) with Dirichlet boundary conditions. In all cases, centered differences were used to discretize the first derivative terms, and the mesh size was $h = 1/32$, so that the order of the linear system was N = 961. The curves in the figures represent the average iteration counts for three test problems, determined by three initial guesses with random values in the interval $[-1, 1]$. In all cases, the right hand side s was identically zero. The convergence criterion was $\|r_i\|_2/\|r_0\|_2 \leq 10^{-6}$, where $r_i = s - Su_i^{(k)} = -Su_i^{(k)}$ is the residual at the i'th iteration.

The left side of each of these figures contains results for the one-line orderings, and the right side contains results for the two-line orderings. Experiments were run for values of γ or δ equal to multiples of 0.2 in $[0, 2]$, plus 7 (or 6) = 3. Fig. 6.1 corresponds to the case $\delta = 0$ (i.e. only the u_x first order term was present in (1.4)), Fig. 6.2 to 7 = 0

[3] This computed spectral radius exceeds the analytic bound. Computations on a Sun 3/60 gave the same results. We believe that this eigenvalue computation is affected by ill-conditioning, although we do not understand the difficulty.

δ	$h = 1/8$	$h = 1/16$	$h = 1/32$	Asymptotic Bound
.2	.42	.74	.85	.90
.4	.32	.54	.62	.65
.6	.19	.30	.34	.36
.8	.07	.11	.12	.12
1.0	0	0	0	0
1.2	.03	.04	.04	
1.4	.06	.08	.09	–
1.6	.09	.13	.13	–
1.8	.12	.16	.18	–
2.0	.14	.20	.22	–
3.0	.21	.36	.41	–

Table 8.2: Spectral **radii and bounds for the two-line Gauss-Seidel iteration** matrices, centered differences, $\gamma = 0$.

γ	$h = 1/8$	$h = 1/16$	$h = 1/32$	Asymptotic **Bound**
.2	.39	.67	.77	.81
.4	.23	.37	.42	.44
.6	.09	.14	.16	.16
.8	.02	.03	.03	.03
1.0	0	0	0	0
1.2	.01	.02	.02	.03
1.4	.04	.05	.05	.13
1.6	.08	.09	.09	.34
1.8	.12	.12	.12	.71
2.0	.16	.16	.16	1.27
3.0	.32	.33	.33	9.00

Table 6.3: Spectral **radii and bounds for the two-line Gauss-Seidel** iteration matrices, centered differences, $\gamma = 6$.

(only u_y), and Fig. 6.3 to $\gamma = \delta$ (u_x and u_y). **The** results are **for the block** Gauss-Seidel **method with the natural, red-black and torus orderings. (The iteration matrices for the alternating torus ordering are similar, via permutation matrices, to those for the one-line red-black ordering, so that these orderings produce identical iterates.) In addition, results for the block SOR method with the natural ordering** are shown **for some choices of γ and 6. For SOR, we used the optimal value of w determined by (3.3), where $\rho(B)^2$ is taken from Tables 6.1 –** 6.3 **and analogous results from [4], using the values far $h = 1/32$.**

We make the following observations concerning these results:

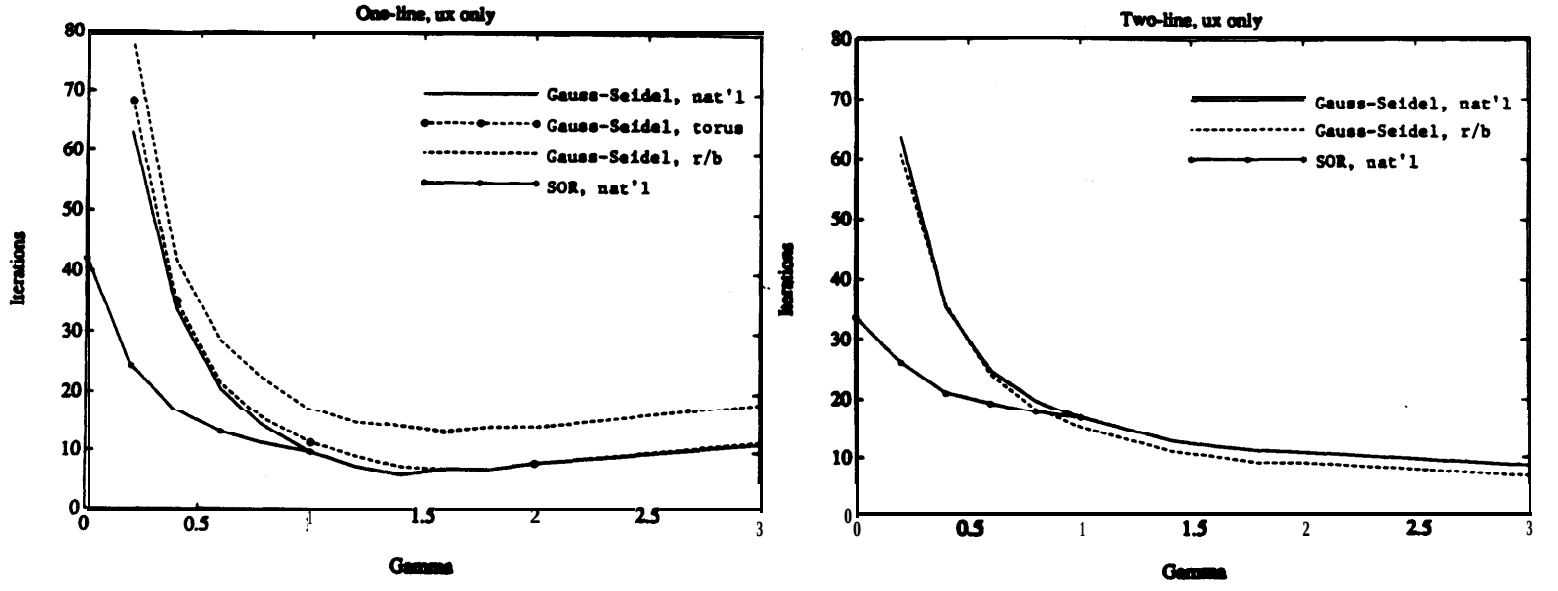

Fig. 6.1: Average iteration counts, $h = 1/32$, $\delta = 0$.

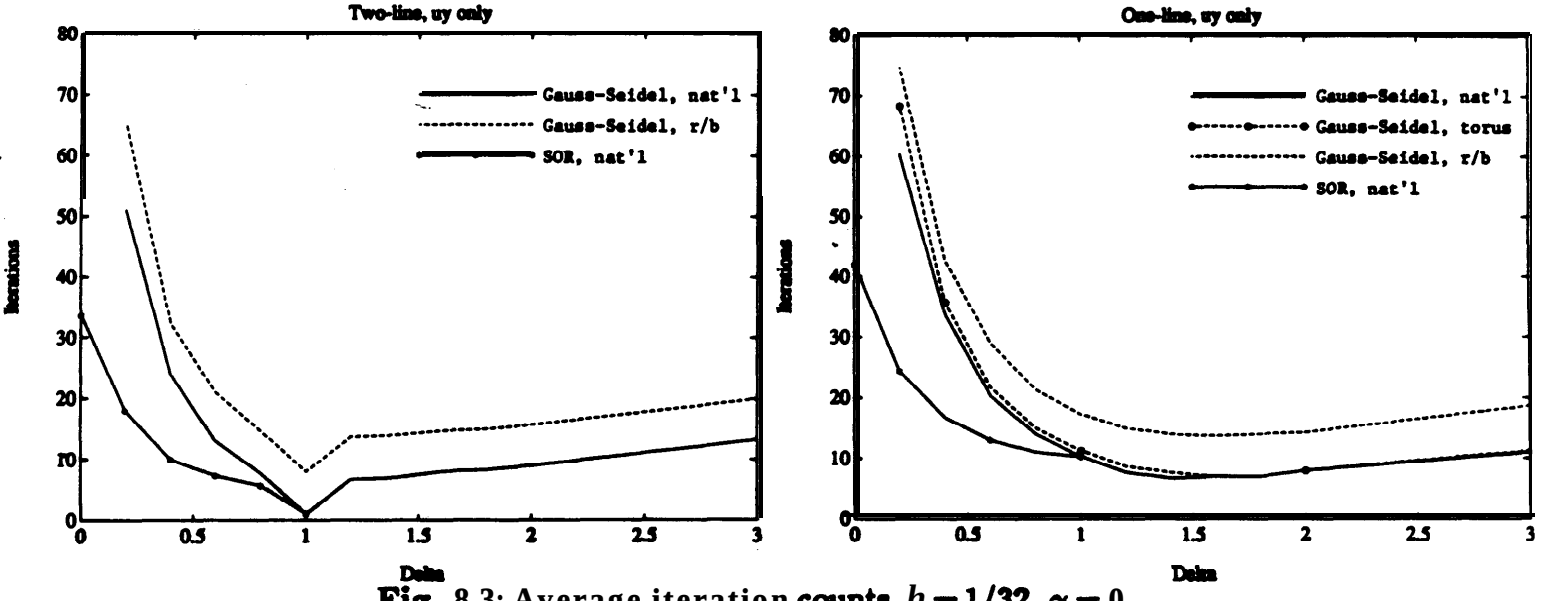

Fig. 8.3: Average iteration counts, $h = 1/32$, $\gamma = 0$.

(1) In most cases, the Gauss-Seidel method requires thirty or fewer iterations to reach the stopping criterion. In general, fewer iterations are required with the natural orderings than with the red-black orderings; a rough estimate is that the red-black orderings entail at most twice as many iterations as the natural orderings. An exception is when $\delta = 0$, where the performances of the natural and red-black two-line orderings are very close (see the right side of Fig. 6.1).

(2) The best results are obtained when γ or δ are near one, and performance typically improves as $|\gamma|$ or $|\delta| \rightarrow 1$. For all values of γ and δ tested, the self-adjoint case ($\gamma = \delta = 0$) required the largest number of Gauss-Seidel iterations. In these cases, for which the results are not shown on the figures, the stopping criterion was typically not reached after 150 iterations. In general, performance is in

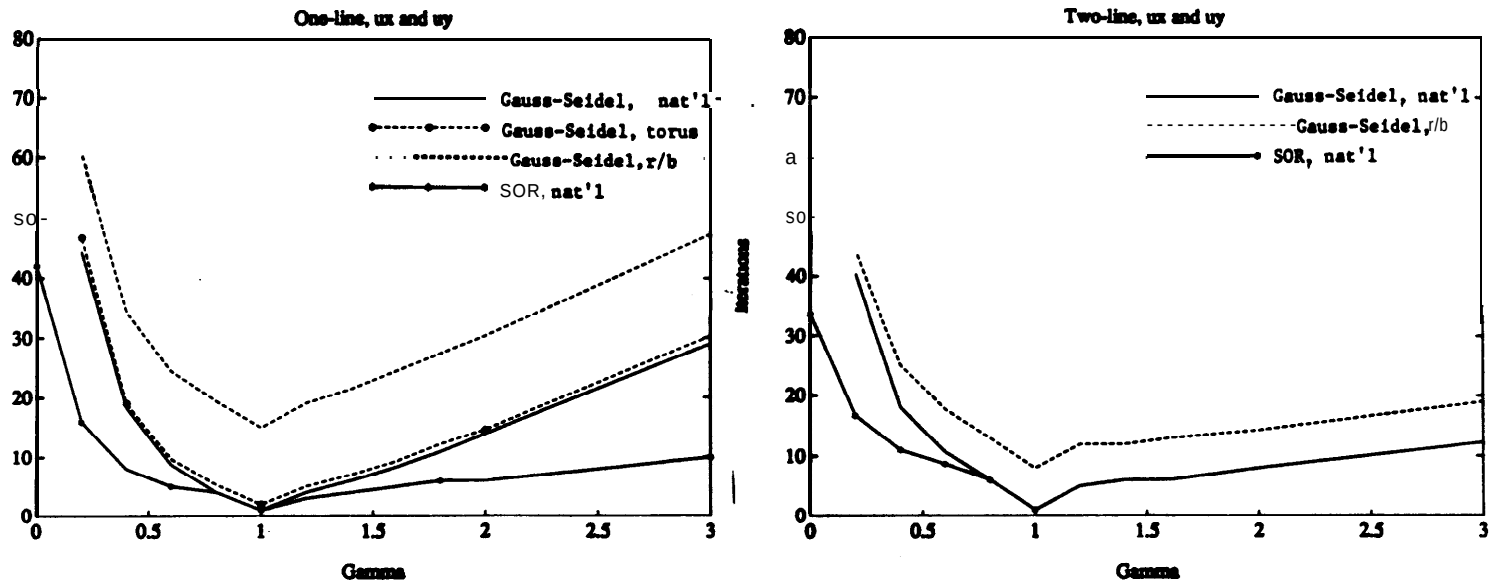

Fig. 6.3: Average iteration counts, $h = 1/32$, $\gamma = 6$.

accordance with the results on spectral radii from Tables 6.1 – 6.3 **and [4].**

(3) **The best results for large** γ **or** δ **are for the two-line orderings with** $\delta = 0$ **(Table** 6.1 **and Fig.** 6.1). **This is because as** $|\gamma|$ **grows, S essentially consists of its block diagonal** D **plus a small perturbation. For large** δ **and** $\gamma = 0$, **a *vertical* two-line splitting would give better results than the horizontal splitting used.**

(4) SOR was much more effective than Gauss-Seidel when the latter was slow. We examined SOR only in cases where the spectrum of the block Jacobi iteration matrix is real, i.e. where either $|\gamma| < 1$ **and** $|\delta| < 1$ **or (for the one-line ordering [4])** $|\gamma| > 1$ **and** $|\delta| > 1$**. Thus, (3.3) applies. In variable coefficient problems of a similar character, it would be realistic to use an adaptive method to estimate the optimal value of w (see e.g. [18]). For other values of** γ **or 6, the spectral radius of the Gauss-Seidel iteration matrix is already very small, and we did not experiment with SOR. To keep the graphs from being too detailed, the SOR results are shown only for the natural orderings. Like Gauss-Seidel, with the red-black orderings SOR typically required somewhat more iterations, but it displayed the same general character as it did with the natural ordering (i.e.** graphs of **iteration counts have similar slopes).**

(5) The performance of the Gauss-Seidel method with the torus ordering is very close to its performance with the natural one-line ordering.

The **error** $e_j \equiv u - u_j$ **at the j'th step of each of the methods under consideration** satisfies $e_j = Me_{j-1}$, **where** M **is the iteration matrix. Thus, for large enough j, the error will be dominated by the eigenvector corresponding to the spectral radius, and the** asymptotic (in terms **of** iteration **counts) analysis of §3 and §4 can be used to predict behavior. However, this does not say anything about how other components of the error affect performance, and it also does not explain the effects of different orderings.** Figures

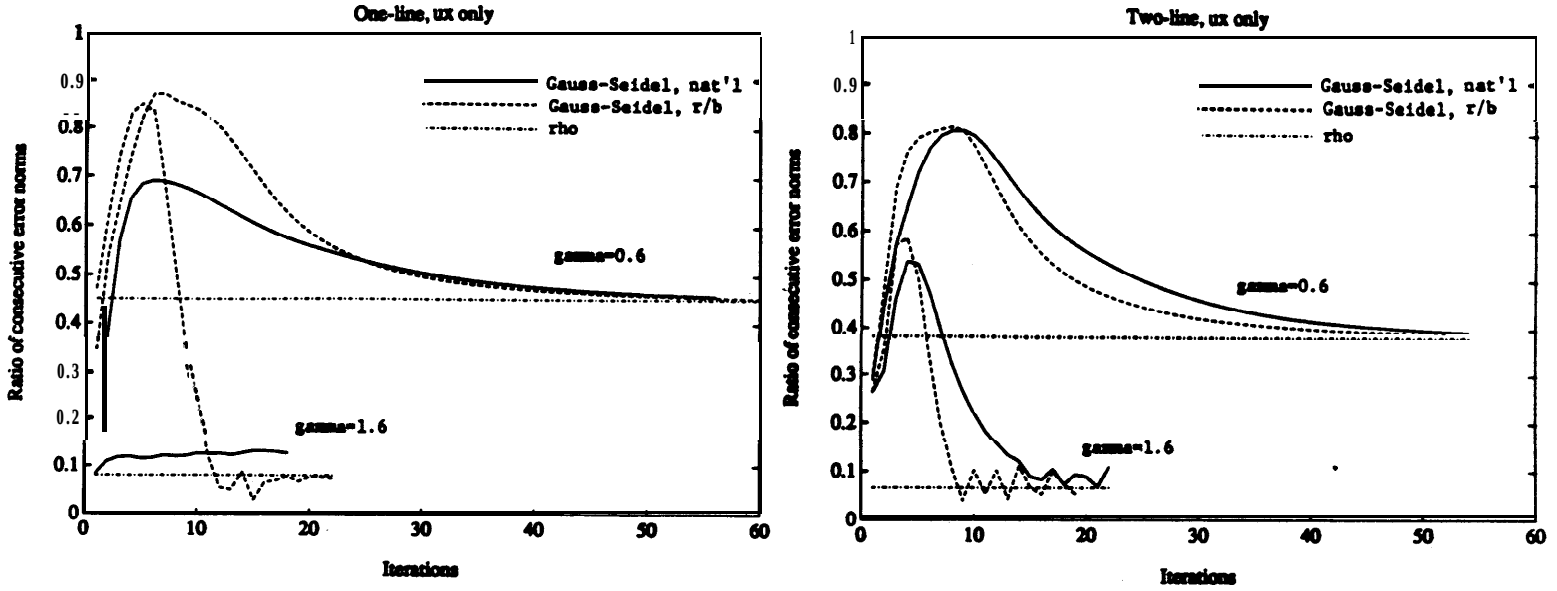

Fig. 6.4: **Approach to asymptotic performance,** $h = 1/32, \delta = 0$.

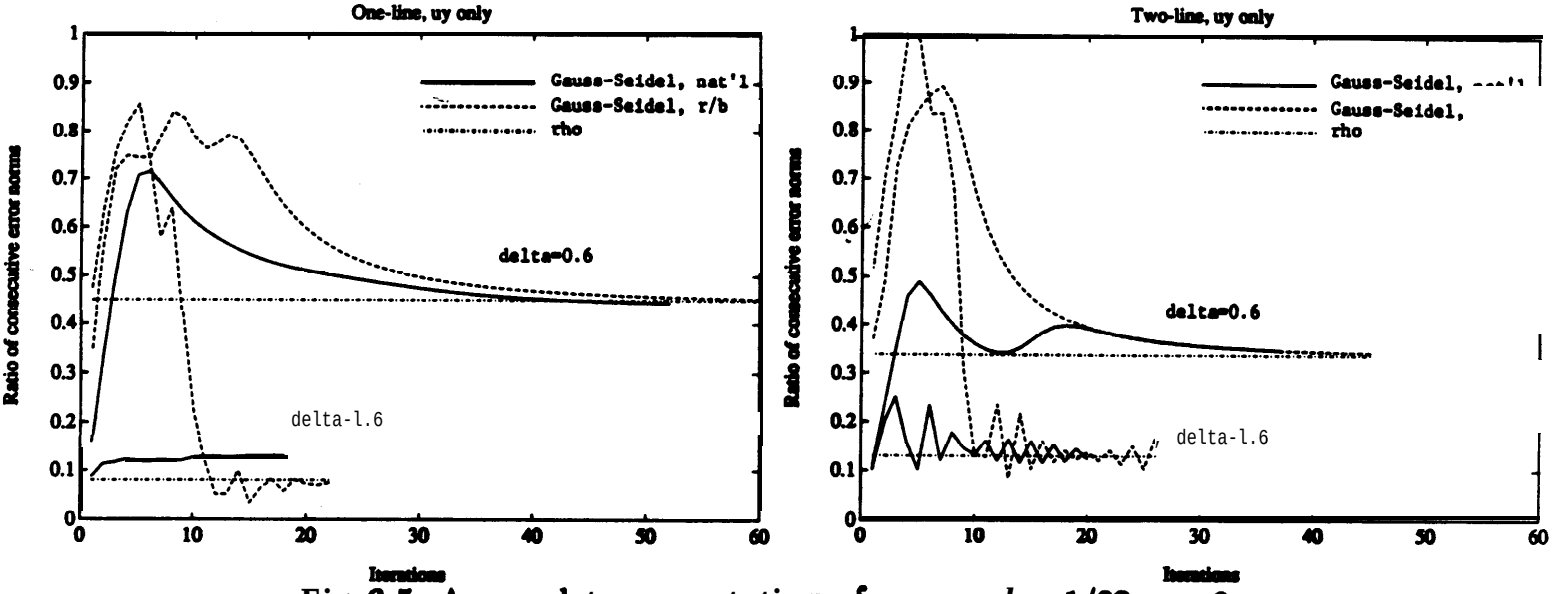

Fig. 6.5: Approach to asymptotic performance, $h = 1/32, \gamma = 0$.

6.4 – 6.6 examine the question of when the asymptotic behavior takes effect, in the Gauss-Seidel method. Each figure graphs the ratio $\|e_j\|_2/\|e_{j-1}\|_2$, **for the natural and red-black versions of both the one-line and two-line orderings, for two problems, one where** 7 **or** δ **is less than one, and one where** 7 **or** δ **is greater than one. Figure** 6.4 **shows the case where** $7 = .6$ **and** 1.6 **and** $\delta = 0$; **Figure** 6.5 **shows the case where** $7 = 0$, **and** $\delta = .6$ **and** 1.6; **and Figure** 6.6 **shows the case where** $7 = \delta = .6$ **and** 1.6. **These results are for one of the initial guesses used in the experiments described above. In all cases, the iterations were performed until the (stringent) stopping criterion** $\|e_i\|_2/\|e_0\|_2 \leq 10^{-16}$ was satisfied.

The results show that the behavior of the Gauss-Seidel method is typically closer to that predicted by the asymptotic analysis when the natural ordering is used, and that fewer iterations are required before the asymptotic performance is seen. The one exception in these examples is Where $\delta = 0$ **with the two-line ordering (Fig. 6.4); in this case the**

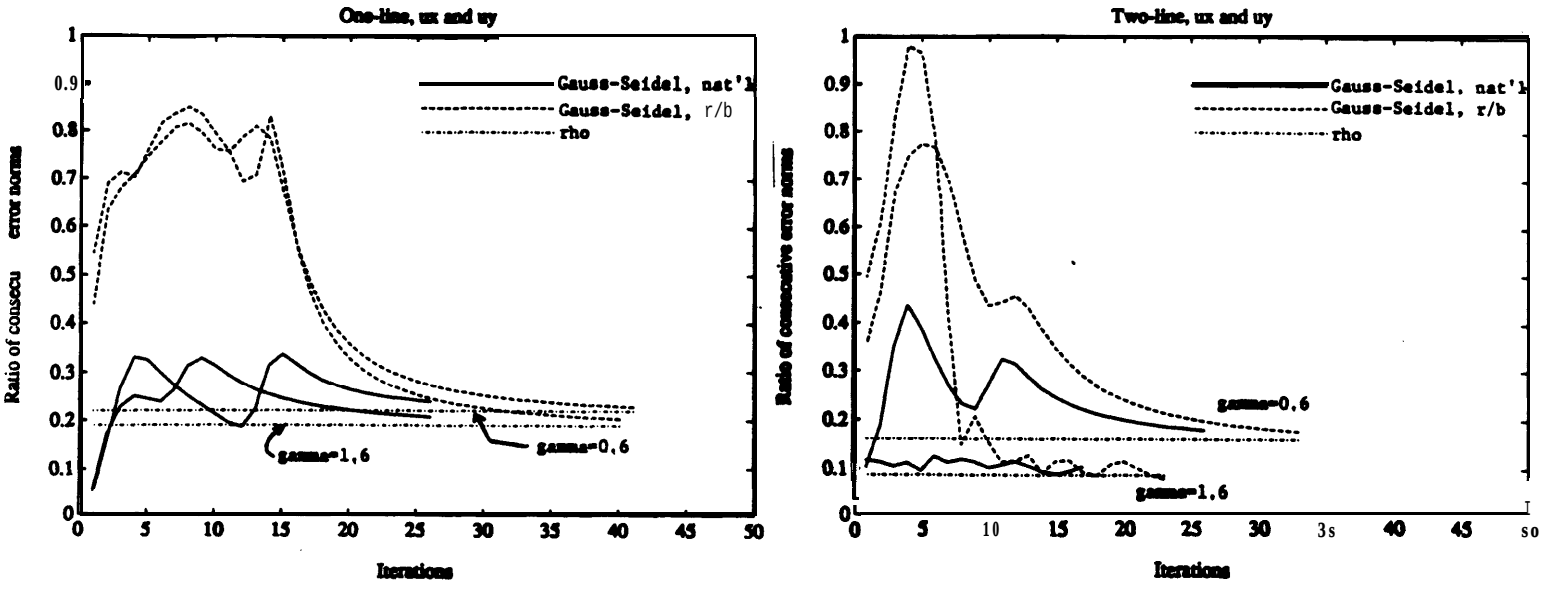

Fig. 6.6: Approach to asymptotic performance, $h = 1/32$, $\gamma = \delta$.

natural and red-black orderings display similar asymptotic behavior. Recall that this was the one case where the performances were similar. We also remark that the asymptotic performance is typically displayed only after the stopping criterion used for Figs. 6.1 – 6.3 is satisfied.

		One-line Natural	One-line R/B	Two-line Natural	Two-line R/B
$\delta = 0$	$\gamma = 0.6$	.86	1.38	1.12	1.35
	$\gamma = 1.6$	.27	1.40	1.00	1.27
$\gamma = 0$	$\delta = 0.6$	.86	1.38	.92	1.47
	$\delta = 1.6$	.27	1.40	1.57	1.65
$\gamma = \delta$	$\gamma = 0.6$	.53	1.40	.87	1.46
	$\gamma = 1.6$	.53	1.40	1.14	1.65

Table 6.4: Euclidean norms of the Gauss-Seidel iteration matrices.

With $M = \mathcal{L}_1$, the errors for the Gauss-Seidel iteration satisfy $e_j = \mathcal{L}_1^j e_0$, so that $\|\mathcal{L}_1^j\|$ would give more precise predictions of the behavior of the errors. Table 6.4 shows $\|\mathcal{L}_1\|_2$ for the twelve examples of Figs. 6.4 – 6.6. These norms were computed by taking the maximum singular value, acquired using LINPACK [3] (in double precision Fortran) on a SUN 3/60. The results show that the norms for the natural orderings are typically less than one, and the norms for the red-black ordering are typically greater than one as well as greater than those for the natural ordering. Thus, the results are largely consistent with the numerical behavior described above. There are cases, however, where $\|\mathcal{L}_1\|_2 > 1$ but the asymptotic behavior is good, e.g. $\gamma = 0$, $\delta = 1.6$, two-line natural ordering (see Fig 6.5).

6.3. **Implementation and parallelism.**

We now outline the implementation costs of the block iterative methods for solving the reduced system. We focus on the block SOR iteration, of which the Gauss-Seidel method is a special case. Assume that the reduced matrix has the form S_{ij}, where the indices refer to the blocks associated with the lines of the ordering in use. For example, for the natural two-line ordering, i and j vary between 1 and $n/2$ and $S_{ij} = 0$ for $|i - j| > 1$. Let $S = D - (L + U)$, where D, L and U are blocked in an analogous manner, and let s and $z \equiv u^{(b)}$ be indiced in an analogous manner. Note that each block of D is a banded matrix of total bandwidth either three (for the one-line orderings) or five (for the two-line orderings). Assume for simplicity that the LU-factorization of each D_i can be computed without pivoting. (This is the case whenever D diagonally dominant.)

The block SOR iteration has the form

$$z_i^{(m+1)} = z_i^{(m)} - \omega\left[z_i^{(m)} - D_i^{-1}\left(\sum_{j<i} L_{ij} z_j^{(m+1)} + \sum_{j>i} U_{ij} z_j^{(m)}\right)\right] + D_i^{-1} s_i, \tag{6.1}$$

where i varies from 1 to the number of blocks in the matrix. Consider the computations involving the matrices D, L and U. Each step requires a matrix-vector product by the i'th block row of U and a matrix-vector product by the i'th block row of L, followed by a linear solve in which the coefficient matrix is the i'th block of D. The cost of the matrix-vector products (in terms of multiply-adds) is essentially. equal to the number of nonzeros in the i'th block rows of L and U. Moreover, assuming that D_i *has* been factored, the cost of the linear solve is equal to the number of nonzeros in D_i. Consequently, for any of the orderings, the total cost of the matrix computations on a serial computer is approximately $9n^2/2$, the number of nonzeros in S. All the other computations (vector adds and scalar-matrix products) are clearly independent of ordering. The factorization of the blocks of D is slightly more expensive for the two-line ordering than for the one-line ordering, but both are of the order of the cost of one iteration, so that the difference is negligible. Pivoting will have a somewhat more detrimental effect on the two-line orderings than on the one-line orderings.

Both the natural and red-black orderings have efficient implementations on parallel computers with k = O(n) processors. The architecture need not have a more complex topology than a linear array (or a ring for the torus orderings), and our discussion applies as well to shared memory machines. It is straightforward to show that the construction of the reduced system is fully parallelizable. In examining the iterative methods, we assume for simplicity that the ordering is such that all block rows of the reduced matrix are of the same size. This is the case for the torus one-line ordering and for the two-line ordering when n is even; the size is approximately n. Let n_r denote the number of block rows; for all orderings, $n_r \approx n/2$. Assume further that k divides n_r, and let the processors be indexed from 1 to k.

The iterations for the natural versions of these orderings can be pipelined using the methods of [1], where a (block) *step* of the computation is defined by the following rule:

> at the i'th step, Processor j is performing the $(i - j + 1)$'st iteration on the first j x n_r/k block rows.

That is, at step one, Processor 1 performs the first iteration on the first n_r/k block rows. Then, at step 2, Processor 2 performs the first iteration on the second n_r/k block rows, and Processor 1 performs the second iteration on the first n_r/k block rows. The first iteration is completed by Processor k after k such steps, and every subsequent step results in the completion of one more iteration. All processors are busy except during the first and last $k - 1$ steps. For t iterations, the speedup (&arithmetic) is

$$\frac{k}{1 + k/t}.$$

Thus, the pipelined implementation is efficient whenever t is large relative to k. For architectures with distributed memory, neighboring processors must exchange vectors of length (approximately) n *between step,* and some overlap of communication and arithmetic is possible.

The alternating torus ordering requires that $\lceil n/2 \rceil$ be even in order to correspond to a red-black ordering; no additional assumptions on n are needed for the two-line red-black ordering. Both red-black orderings are then fully parallelizable on up to $n/4$ processors. For all indices i with red color, the computation (6.1) consists of a set of independent block matrix-vector products by the nonzero blocks of U, followed by a set of independent block matrix solves. Then, for all indices i with black color, the steps of (6.1) consist of a set of independent block matrix-vector products by the nonzero blocks of L, followed by a set of independent block matrix solves. Unidirectional communication between neighboring processors of vectors of length n is needed twice, prior to the multiplications by L and U. Overlap with arithmetic is possible.

In §6.2, we found the methods to be very effective on model problems. For the small values of t observed, it appears that the inefficiency of the natural orderings due to pipelining will often be similar in scale to the somewhat slower performance displayed by the red-black orderings. Consequently, we expect the performance of the two classes of orderings to be comparable on parallel architectures.

7. Concluding remarks.

In this paper, we have continued the analysis begun in [4] of block iterative methods for solving cyclically reduced linear systems derived from the convection-diffusion equation. We showed how the discrete reduced system is related to the underlying continuous problem, and we derived bounds on the spectral radius of the block Jacobi iteration matrix associated with two-line orderings of the reduced grid. These bounds, together with analogous ones from [4], were combined with the Young theory to analyze the asymptotic convergence behavior of the Gauss-Seidel and SOR block iterative methods derived from several variants of both two-line orderings and one-line orderings. The results express convergence behavior in terms of discrete cell Reynolds numbers $\sigma h/2, \tau h/2$, and they are confirmed and supplemented by numerical experiments. The analytic and experimental results (as well as those of [2] and [14]) show that the nonsymmetric discrete problems arising from (1.4) are in some ways easier to solve than the symmetric ones.

Acknowledgements: The authors wish to thank Dianne O'Leary and Seymour Parter for some helpful remarks.

References

[1] L. M. Adams and H. F. Jordan, Is SOR color blind? *SIAM J. Sci. Stat. Comput.* 7:490-506, 1986.

[2] R. C. Y. Chin and T. A. Manteuffel, An analysis of block successive overrelaxation for *a* class of matrices with complex spectra, *SIAM J. Numer. Anal.* 25:564-585, 1988.

[3] J. J. Dongarra, J. R. Bunch, C. B. Moler and G. W. Stewart, *LINPACK Users' Guide*, SIAM Publications, Philadelphia, 1979.

[4] H. C. Elman and G. H. Golub, Iterative Methods for Cyclically Reduced Non-Self-Adjoint Linear Systems, UMIACS Report 88-87, Univ. of Maryland, 1988. To appear in *Math. Comp.*

[5] G. H. Golub and R. S. Varga, Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods, *Numer. Math.* 3:147-156, 1961.

[6] L. A. Hageman and R. S. Varga, Block iterative methods for cyclically reduced matrix equations, *Numer. Math.* 6:106-119, 1964.

[7] R. S. Garbow, J. M. Boyle, J. J. Dongarra, and C. B. Moler, *Matrix Eigensystem Routines: EISPACK Guide Extension*, Springer-Verlag, New York, 1972.

[8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1983.

[9] MACSYMA Reference Manual, Laboratory for Computer Science, MIT, 1977.

[10] T. A. Manteuffel, Optimal parameters for linear seconddegree stationary iterative methods, *SIAM J. Numer. Anal.* 833-839, 1952.

[11] S. V. Parter, Iterative Methods for elliptic problems and the discovery of "q," *SIAM Reo.* 28:153-175, 1986.

[12] S. V. Parter, On estimating the "rates of convergence" of iterative methods for elliptic difference equations, *Trans. Amer. Math. Soc.* 114:320-354, 1965.

[13] S. V. Parter, On "two-line" iterative methods far the Laplace and biharmonic difference equations, *Numer. Math.* 1:240-252, 1959.

[14] S. V. Parter and M. Steuerwalt, Block iterative methods for elliptic and parabolic difference equations, *SIAM J. Numer. Anal.* 19:1173-1195, 1982.

[15] PCGPAK User's Guide, Version 1.04, Scientific Computing Associates, New Haven, CT, 1987.

[16] A. Segal, Aspects of numerical methods for elliptic singular perturbation problems, *SIAM J. Sci. Stat. Comput.* 3:327-349, 1982.

[17] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, New Jersey, 1962.

[18] D. M. Young, *Iterative Solution of Large Linear System*, Academic Press, New York, 1971.